



Lexicale convergentie in de Nederlandse woordenschat

Jocelyne Daems Kris Heylen Dirk Geeraerts



KU Leuven

OG Kwantitatieve Lexicologie en Variatielinguïstiek

Overzicht

1. Inleiding

2. Het Nederlands

3. Methode

Uniformiteitsmaat

Data

4. Resultaten

5. Conclusie

Referenties



1. Inleiding

Variatie binnen het Nederlands

Lectale variatie

- register
- regio
- geslacht
- generatie
- beroep en opleiding

Talige variatie

- vrachtwagen of vrachtauto (of camion of truck)
- dat ken/kan/kun jij wel zeggen
- zeven of zeuven



1. Inleiding

Variatie binnen het Nederlands

Lectale variatie

- register
- regio
- geslacht
- generatie
- beroep en opleiding

Een specifieke context kan hand in hand gaan met een groep van typische talige keuzes
→ lect

Talige variatie

- vrachtwagen of vrachtauto
- dat ken/kan jij wel zeggen
- zeven of zeuven



1. Inleiding

Variatie binnen het Nederlands *in deze studie*

Lectale variatie

- regio en register

⇒ geografische en stilistisch/stratificationele vraagstelling

Talige variatie

- lexicon

⇒ onomasiologische vraagstelling



1. Inleiding

Onderzoeksopzet

Lexicale convergentie meten tussen Nederlands uit België en Nederlands uit Nederland
a.d.h.v. de woordkeuze voor emotie-, ICT- en verkeersconcepten.
⇒ Cf. Geeraerts, Grondelaers & Spielman 1999



Overzicht

1. Inleiding

2. Het Nederlands

3. Methode

4. Resultaten

5. Conclusie

Referenties



2. Het Nederlands

Het? Nederlands

Twee nationale variëteiten (pluricentrische taal (Clyne 1992))

- Nederlands Nederlands
- Belgisch Nederlands

Stratificatie

- Poldernederlands ?
- Verkavelingsvlaams, soapvlaams, tussentaal



2. Het Nederlands

Standaardisatie

Verschillend verloop standaardisatieproces

- Nederland: onafhankelijke ontwikkeling van een (de?) standaardvariant van het Nederlands
- België: na een langdurige belemmering, bevordering van de convergentie met de (gevestigde) Nederlands-Nederlandse norm

Twee tegengestelde perspectieven in België

- Integrationisme: Nederlands Nederlands als de norm
- Particularisme: ruimte voor een Belgisch-Nederlandse standaardvariant



Overzicht

1. Inleiding

2. Het Nederlands

3. **Method**

Uniformiteitsmaat

Data

4. Resultaten

5. Conclusie

Referenties



3. Methode

Geeraerts, Grondelaers & Speelman 1999

- uniformiteitsmaat voor onomasiologische variatie

3. Methode

Geeraerts, Grondelaers & Speelman 1999

- uniformiteitsmaat voor onomasiologische variatie
- ⇒ in hoeverre worden voor een bepaald concept dezelfde woorden gekozen?

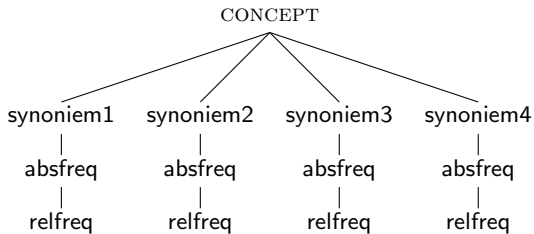
3. Methode

Geeraerts, Grondelaers & Spielman 1999

- uniformiteitsmaat voor onomasiologische variatie
- ⇒ in hoeverre worden voor een bepaald concept dezelfde woorden gekozen?
- ⇒ hoe vaak drukken Belgen en Nederlanders VRACHTWAGEN hetzelfde uit?

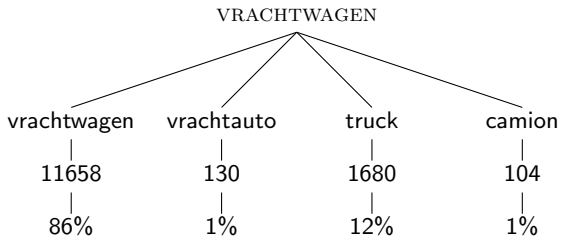
3. Methode - Uniformiteitsmaat

Onomasiologisch profiel



3. Methode - Uniformiteitsmaat

Onomasiologisch profiel: concreet



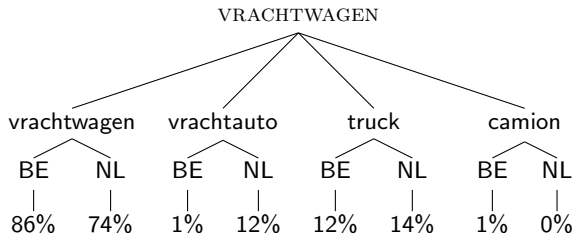
3. Methode - Uniformiteitsmaat

Lexicale convergentie meten: vijf stappen

1. Per concept het onomasiologisch profiel opstellen
2. Per variëteit het onomasiologisch profiel opbouwen
3. Uniformiteit meten
4. Aggregeren over alle concepten
5. (De uniformiteitsmaat wegen d.m.v. de conceptfrequentie)

3. Methode - Uniformiteitsmaat

Lexicale convergentie meten: concreet



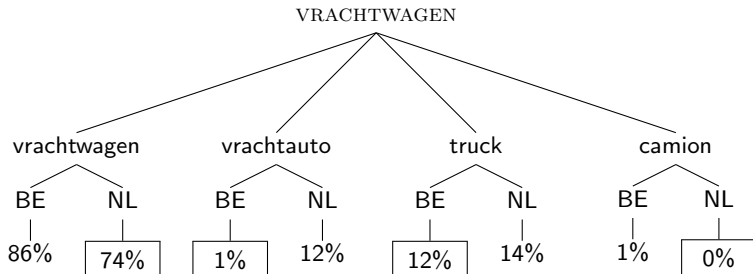
3. Methode - Uniformiteitsmaat

Lexicale convergentie meten: vijf stappen

1. Per concept het onomasiologisch profiel opstellen
2. Per variëteit het onomasiologisch profiel opbouwen
3. Uniformiteit meten
 - ⇒ = overlapping in lexicaliseringsvoorkeuren
 - = som van de kleinste relatieve waarde voor elke term
4. Aggregeren over alle concepten
5. (De uniformiteitsmaat wegen d.m.v. de conceptfrequentie)

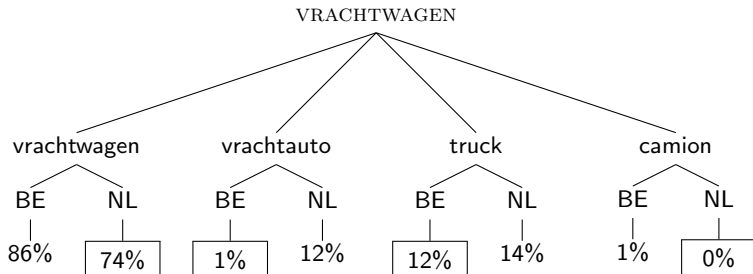
3. Methode - Uniformiteitsmaat

Lexicale convergentie meten: concreet



3. Methode - Uniformiteitsmaat

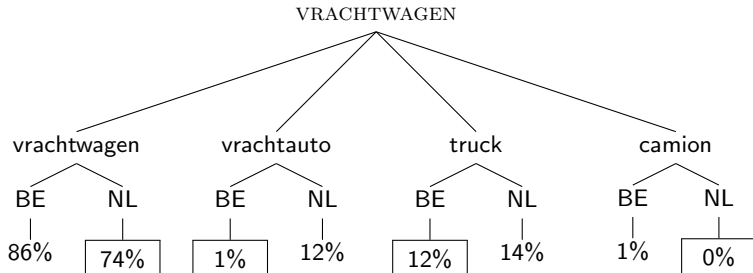
Lexicale convergentie meten: concreet



$$\Rightarrow 74\% + 1\% + 12\% + 0\% = 87\%$$

3. Methode - Uniformiteitsmaat

Lexicale convergentie meten: concreet



$\Rightarrow 74\% + 1\% + 12\% + 0\% = 87\%$ \leftarrow mate van uniformiteit

3. Methode - Uniformiteitsmaat

Lexicale convergentie meten: vijf stappen

1. Per concept het onomasiologisch profiel opstellen
2. Per variëteit het onomasiologisch profiel opbouwen
3. Uniformiteit meten
4. Aggregeren over alle concepten
⇒ = herhalen voor ieder concept in het semantisch veld
+ het gemiddelde berekenen
5. (De uniformiteitsmaat wegen d.m.v. de conceptfrequentie)



3. Methode - Uniformiteitsmaat

Lexicale convergentie meten: vijf stappen

	U
VRACHTWAGEN	87%
METRO	97%
TREIN	99%
STOEP	61%
BESTELWAGEN	45%
gemiddelde	78%



3. Methode - Uniformiteitsmaat

Lexicale convergentie meten: vijf stappen

	U
VRACHTWAGEN	87%
METRO	97%
TREIN	99%
STOEP	61%
BESTELWAGEN	45%
gemiddelde	78%

3. Methode - Uniformiteitsmaat

Lexicale convergentie meten: vijf stappen

1. Per concept het onomasiologisch profiel opstellen
2. Per variëteit het onomasiologisch profiel opbouwen
3. Uniformiteit meten
4. Aggregeren over alle concepten
5. De uniformiteitsmaat wegen d.m.v. de conceptfrequentie
⇒ = minder frequente concepten tellen minder mee

3. Methode - Uniformiteitsmaat

Lexicale convergentie meten: vijf stappen

	U
VRACHTWAGEN	87%
METRO	97%
TREIN	99%
STOEP	61%
BESTELWAGEN	45%
gemiddelde	78%

3. Methode - Uniformiteitsmaat

Lexicale convergentie meten: vijf stappen

	U	W
VRACHTWAGEN	87%	32%
METRO	97%	8%
TREIN	99%	46%
STOEP	61%	10%
BESTELWAGEN	45%	4%
gemiddelde	78%	

3. Methode - Uniformiteitsmaat

Lexicale convergentie meten: vijf stappen

	U	W	U'
VRACHTWAGEN	87%	32%	28%
METRO	97%	8%	7%
TREIN	99%	46%	45%
STOEP	61%	10%	6%
BESTELWAGEN	45%	4%	2%
gemiddelde	78%	som	89%

3. Methode - Data

Dataverzameling

Vier stappen

- Selectie profielen
- Gelemmatiseerde corpora
- Script frequenties en concordanties
- Manuele desambiguatie

3. Methode - Data

Dataverzameling

Vier stappen

- Selectie profielen
- Gelemmatiseerde corpora
- Script frequenties en concordanties
- Manuele desambiguatie



3. Methode - Data

Profielen verkeer



rondpunt
rotonde



pechstrook
vluchstrook



voetgangersoversteekplaats
zebra zebraad

3. Methode - Data

Profielen verkeer



convertible
cabrio
cabriolet



fiets
rijwiel
velo
tweewieler



rondpunt
rotonde



pechstrook
vluchstrook



ondergrondse
metro
metrotrein



voetgangersoversteekplaats
zebra zebraapad



3. Methode - Data

Profielen ICT



mobiel gsm
mobieltje
mobiele telefoon



laptop
draagbare
computer
schootcomputer
notebook



klavier
keyboard
toetsenbord

3. Methode - Data

Profielen ICT



mobiel gsm
mobieltje
mobiele telefoon



desktop
bureaublad



laptop
draagbare
computer
schootcomputer
notebook



klavier
keyboard
toetsenbord



wachtwoord
paswoord
password

3. Methode - Data

Profielen emoties

kolerig
boos

schrik
angst
vrees

degout
afgrijzen

razend
kwaad
woedend



afschuw
weezin
afkeer

verbazing



droevig
triest
verdrietig

verstomming

blij vrolijk
opgetogen



3. Methode - Data

Dataverzameling

Vier stappen

- Selectie profielen
- Gelemmatiseerde corpora
- Script frequenties en concordanties
- Manuele desambiguatie



3. Methode - Data

Een stukje geannoteerde tekst door Alpino

['Na', 'prep', 'na', 3849, 0], ['een', 'det', 'een', 3849, 1], ['jaar', 'noun', 'jaar', 3849, 2],
 ['- ', 'punct', '-', 3849, 3], ['ik', 'pron', 'ik', 3849, 4], ['deed', 'verb', 'doe', 3849, 5],
 ['alles', 'noun', 'alles', 3849, 6], [',', 'punct', ',', 3849, 7], ['van', 'prep', 'van', 3849, 8],
 ['was', 'noun', 'was', 3849, 9], ['ophangen', 'verb', 'hang_op', 3849, 10], ['tot', 'prep',
 'tot', 3849, 11], ['de', 'det', 'de', 3849, 12], ['beenhouwerij', 'noun', 'beenhouwerij',
 3849, 13], ['kuisen', 'verb', 'kuis', 3849, 14], ['- ', 'punct', '-', 3849, 15], ['vroeg',
 'verb', 'vraag', 3849, 16], ['ik', 'pron', 'ik', 3849, 17], ['of', 'comp', 'of', 3849, 18],
 ['ik', 'pron', 'ik', 3849, 19], ['meer', 'det', 'meer', 3849, 20], ['kon', 'verb', 'kan', 3849,
 21], ['verdienen', 'verb', 'verdien', 3849, 22], [',', 'punct', ',', 3849, 23], ['De', 'det',
 'de', 3850, 0], ['slager', 'noun', 'slager', 3850, 1], ['wees', 'verb', 'wijs_af', 3850, 2],
 ['dat', 'det', 'dat', 3850, 3], ['verzoek', 'noun', 'verzoek', 3850, 4], ['af', 'part', 'af',
 3850, 5], [',', 'punct', ',', 3850, 6], ['Ik', 'pron', 'ik', 3851, 1], ['pakte', 'verb', 'pak',
 3851, 2], ['m'n', 'det', 'mijn', 3851, 3], ['velo', 'noun', 'velo', 3851, 4], ['en', 'vg',
 'en', 3851, 5], ['ik', 'pron', 'ik', 3851, 6], ['was', 'verb', 'ben', 3851, 7], ['weg', 'adv',
 'weg', 3851, 8], [',', 'punct', ',', 3851, 9]

⇒ Na een jaar – ik deed alles, van was ophangen tot de beenhouwerij kuisen – vroeg ik of ik meer kon verdienen. De slager wees dat verzoek af. Ik pakte m'n **velo** en ik was weg.



3. Methode - Data

Dataverzameling

Vier stappen

- Selectie profielen
- Gelemmatiseerde corpora
- Script frequenties en concordanties
- Manuele desambiguatie



3. Methode - Data

Dataverzameling

Vier stappen

- Selectie profielen
- Gelemmatiseerde corpora
- Script frequenties en concordanties
- Manuele desambiguatie



3. Methode - Data

Manuele desambiguatie

Stevaert vindt de sluiting van de kleine **ring** rondom Antwerpen een topprioriteit.

Al draagt een aap een gouden **ring**, het is en blijft een lelijk ding.

Wat is de kick ? Als ik in de **ring** oog in oog sta met mijn tegenstander.



3. Methode - Data

Manuele desambiguatie

Stevaert vindt de sluiting van de kleine **ring** rondom Antwerpen een topprioriteit. ⇒ **relevante betekenis**

Al draagt een aap een gouden **ring**, het is en blijft een lelijk ding.
⇒ **verkeerde betekenis: sieraad**

Wat is de kick ? Als ik in de **ring** oog in oog sta met mijn tegenstander. ⇒ **verkeerde betekenis: boksring**

3. Methode - Data

Profielen

80 concepten

2 woordsoorten

3 lexicale velden

- verkeer: VRACHTWAGEN
- ICT: GSM
- emotie: VERLEGEN, OPRECHTHEID

Corpora

	kwal.krant	Usenet
België	373M	18M
Nederland	161M	24M



Overzicht

1. Inleiding

2. Het Nederlands

3. Methode

4. Resultaten

5. Conclusie

Referenties



4. Resultaten

Hypotheses

1. Diachrone convergentie
= het woordgebruik in België en Nederland wordt uniformer
2. Synchrone stratificatie
= het woordgebruik in Nederland is uniformer dan in België
= het woordgebruik in kranten is uniformer dan in Usenet
3. Rol van het semantische veld
4. Rol van de conceptfrequentie



4. Resultaten

Hypothese 1: Diachrone convergentie

Studie : Geeraerts, Grondelaers & Speelman 1999

Materiaal : kledingtermen in modemagazines in 1990

$U'_{(B50,N50)}$ 70%

$U'_{(B70,N70)}$ 75%

$U'_{(B90,N90)}$ 82%

4. Resultaten

Hypothese 1: Diachrone convergentie

Studie : Geeraerts, Grondelaers & Speelman 1999

Materiaal : kledingtermen in modemagazines in 1990

$U'_{(B50,N50)}$ 70%

$U'_{(B70,N70)}$ 75%

$U'_{(B90,N90)}$ 82%

Studie : De Cnodder 2013

Materiaal : kledingtermen in modemagazines in 2012

$U'_{(B12,N12)}$ 78%

4. Resultaten

Hypothese 1: Diachrone convergentie

Studie : Geeraerts, Grondelaers & Spielman 1999

Materiaal : voetbaltermen in tijdschriften in 1990

$U'_{(B50,N50)}$ 66%

$U'_{(B70,N70)}$ 72%

$U'_{(B90,N90)}$ 77%

4. Resultaten

Hypothese 1: Diachrone convergentie

Studie : Grondelaers, Van Aken, Speelman & Geeraerts 2001

Materiaal : kledingtermen in kranten in 1998

$U'_{(B58,N58)}$ 71%

$U'_{(B78,N78)}$ 64%

$U'_{(B98,N98)}$ 82%

4. Resultaten

Hypothese 1: Diachrone convergentie

Studie : Grondelaers, Van Aken, Speelman & Geeraerts 2001

Materiaal : kledingtermen in kranten in 1998

$U'_{(B58,N58)}$ 71%

$U'_{(B78,N78)}$ 64%

$U'_{(B98,N98)}$ 82%

Studie : Grondelaers, Van Aken, Speelman & Geeraerts 2001

Materiaal : voorzetsels in kranten in 1998

$U'_{(B58,N58)}$ 59%

$U'_{(B78,N78)}$ 62%

$U'_{(B98,N98)}$ 67%

4. Resultaten

Hypothese 1: Diachrone convergentie

- Convergentie bij supraregionaal/formeel taalgebruik
50 – 70 – 90 / 58 – 78 – 98
- Verschuiving binnen het Belgisch Nederlands, dat convergeert naar het Nederlands Nederlands

4. Resultaten

Hypothese 1: Diachrone convergentie

- Convergentie bij supraregionaal/formeel taalgebruik
50 – 70 – 90 / 58 – 78 – 98
- Verschuiving binnen het Belgisch Nederlands, dat convergeert naar het Nederlands Nederlands

! De situatie in 2012: stagnatie of divergentie?

- Informalisering Nederlands Nederlands?

! Gebalanceerde samenstelling van het corpus

! Inhoudswoorden vs. functiewoorden

4. Resultaten

Hypothese 2: Synchrone stratificatie

Studie : Geeraerts, Grondelaers & Speelman 1999

Materiaal : kledingtermen in modemagazines en etalages in 1990

$$U'(B_{mag}, B_{etal}) \quad 50\%$$

$$U'(N_{mag}, N_{etal}) \quad 69\%$$

4. Resultaten

Hypothese 2: Synchronie stratificatie

Studie : Geeraerts, Grondelaers & Speelman 1999

Materiaal : kledingtermen in modemagazines en etalages in 1990

$$U'(B_{mag}, B_{etal}) \quad 50\%$$

$$U'(N_{mag}, N_{etal}) \quad 69\%$$

Studie : De Cnodder 2013

Materiaal : kledingtermen in modemagazines en etalages in 2012

$$U'(B_{mag}, B_{etal}) \quad 63\%$$

$$U'(N_{mag}, N_{etal}) \quad 79\%$$

4. Resultaten

Hypothese 2: Synchronie stratificatie

Studie : Grondelaers, Van Aken, Speelman & Geeraerts 2001

Materiaal : kledingtermen in kranten, Usenet en IRC in 1998

$$U'(B_{krant}, B_{Usenet}) \quad 90\%$$

$$U'(B_{krant}, B_{IRC}) \quad 83\%$$

$$U'(N_{krant}, N_{Usenet}) \quad 92\%$$

$$U'(N_{krant}, N_{IRC}) \quad 90\%$$

4. Resultaten

Hypothese 2: Synchronie stratificatie

Studie : Grondelaers, Van Aken, Speelman & Geeraerts 2001

Materiaal : kledingtermen in kranten, Usenet en IRC in 1998

$$U'(B_{krant}, B_{Usenet}) \quad 90\%$$

$$U'(B_{krant}, B_{IRC}) \quad 83\%$$

$$U'(N_{krant}, N_{Usenet}) \quad 92\%$$

$$U'(N_{krant}, N_{IRC}) \quad 90\%$$

Studie : Grondelaers, Van Aken, Speelman & Geeraerts 2001

Materiaal : voorzetsels in kranten, Usenet en IRC in 1998

$$U'(B_{krant}, B_{Usenet}) \quad 82\%$$

$$U'(B_{krant}, B_{IRC}) \quad 75\%$$

$$U'(N_{krant}, N_{Usenet}) \quad 97\%$$

$$U'(N_{krant}, N_{IRC}) \quad 94\%$$

4. Resultaten

Hypothese 2: Synchronie stratificatie

Studie : Daems, Heylen & Geeraerts (In prep.)

Materiaal : verkeerstermen in kranten en Usenet in 1999-2005

$$U_{(B_{krant}, B_{Usenet})} \quad 87\%$$

$$U_{(N_{krant}, N_{Usenet})} \quad 86\%$$

Studie : Daems, Heylen & Geeraerts (In prep.)

Materiaal : ICT-termen in kranten en Usenet in 1999-2005

$$U_{(B_{krant}, B_{Usenet})} \quad 72\%$$

$$U_{(N_{krant}, N_{Usenet})} \quad 70\%$$

4. Resultaten

Hypothese 2: Synchronie stratificatie

Studie : Daems, Heylen & Geeraerts (In prep.)

Materiaal : verkeerstermen in kranten en Usenet in 1999-2005

$U_{(B_{krant}, B_{Usenet})}$ 87%

$U_{(N_{krant}, N_{Usenet})}$ 86%

Studie : Daems, Heylen & Geeraerts (In prep.)

Materiaal : ICT-termen in kranten en Usenet in 1999-2005

$U_{(B_{krant}, B_{Usenet})}$ 72% \Rightarrow *mailinglist, provider, database*

$U_{(N_{krant}, N_{Usenet})}$ 70% \Rightarrow *mailinglist, provider, password*

4. Resultaten

Hypothese 2: Synchronie stratificatie

- Binnen het Belgisch Nederlands ligt de uniformiteit lager
- Binnen het Belgisch Nederlands is er een continuüm van regionaal/informeel (IRC) naar supraregionaal/formeel (kranten) taalgebruik waarbij Usenet een intermediaire positie inneemt

! Invloed semantisch veld

4. Resultaten

Hypothese 2: Synchronie stratificatie (bis)

Studies : Geeraerts, Grondelaers & Speelman 1999,
Grondelaers, Van Aken, Speelman & Geeraerts 2001

Materiaal : kledingtermen in modemagazines in 1990 en 2012,
kranten, Usenet en IRC in 1998

$$U'(B_{mag,90}, N_{mag,90}) \quad 82\%$$

$$U'(B_{mag,12}, N_{mag,12}) \quad 78\%$$

$$U'(B_{krant}, N_{krant}) \quad 82\%$$

$$U'(B_{Usenet}, N_{Usenet}) \quad 73\%$$

$$U'(B_{IRC}, N_{IRC}) \quad 75\%$$

4. Resultaten

Hypothese 2: Synchronie stratificatie (bis)

Studie : Daems, Heylen & Geeraerts (In prep.)

Materiaal : verkeerstermen in kranten en Usenet in 1999-2005

$$U'(B_{krant}, N_{krant}) \quad 78\%$$

$$U'(B_{Usenet}, N_{Usenet}) \quad 72\%$$

Studie : Daems, Heylen & Geeraerts (In prep.)

Materiaal : ICT-termen in kranten en Usenet in 1999-2005

$$U'(B_{krant}, N_{krant}) \quad 74\%$$

$$U'(B_{Usenet}, N_{Usenet}) \quad 91\%$$

4. Resultaten

Hypothese 2: Synchronie stratificatie (bis)

Studie : Daems, Heylen & Geeraerts (In prep.)

Materiaal : verkeerstermen in kranten en Usenet in 1999-2005

$U'(B_{krant}, N_{krant})$ 78%

$U'(B_{Usenet}, N_{Usenet})$ 72% \Rightarrow BUS, TREIN, VRACHTWAGEN

Studie : Daems, Heylen & Geeraerts (In prep.)

Materiaal : ICT-termen in kranten en Usenet in 1999-2005

$U'(B_{krant}, N_{krant})$ 74%

$U'(B_{Usenet}, N_{Usenet})$ 91% \Rightarrow WEBSITE, LINK, PROVIDER

4. Resultaten

Hypothese 2: Synchronie stratificatie (bis)

- Op supraregionaal vlak is de standaardvariant uniformer dan de informele variant(en)

! Invloed semantisch veld



4. Resultaten

Hypothese 3: Rol van het semantische veld

Studies : Geeraerts, Grondelaers & Speelman 1999,
Grondelaers, Van Aken, Speelman & Geeraerts 2001

Materiaal : voetbaltermen in tijdschriften in 1990,
voorzetsels en kledingtermen in kranten in 1998

$U'(B_{kleding}, N_{kleding})$ 82%

$U'(\text{voorzetsel}, N_{\text{voorzetsel}})$ 67%

$U'(B_{voetbal}, N_{voetbal})$ 77%

4. Resultaten

Hypothese 3: Rol van het semantische veld

Studie : Daems, Heylen & Geeraerts (In prep.)

Materiaal : verkeers-, ICT- en emotietermen in kranten in 1999-2005

U' ($B_{verkeer}, N_{verkeer}$)	78%
U' (B_{ICT}, N_{ICT})	74%
U' ($B_{emotie, noun}, N_{emotie, noun}$)	81%
U' ($B_{emotie, adj}, N_{emotie, adj}$)	79%

4. Resultaten

Hypothese 3: Rol van het semantische veld

- Leeftijd van het veld, mate van supraregionaal contact, register

! Statistische significantie: voldoende data nodig

4. Resultaten

Hypothese 4: Rol van de conceptfrequentie

Studie : Daems, Heylen & Geeraerts (In prep.)

Materiaal : verkeers-, ICT- en emotietermen in kranten in 1999-2005

	niet-gewogen (U)	gewogen (U')
$(B_{verkeer}, N_{verkeer})$	70%	78%
(B_{ICT}, N_{ICT})	79%	74%
$(B_{emotie, noun}, N_{emotie, noun})$	83%	81%
$(B_{emotie, adj}, N_{emotie, adj})$	78%	79%

4. Resultaten

Hypothese 4: Rol van de conceptfrequentie

Studie : Daems, Heylen & Geeraerts (In prep.)

Materiaal : verkeers-, ICT- en emotietermen in kranten in 1999-2005

	niet-gewogen (U)	gewogen (U')
$(B_{verkeer}, N_{verkeer})$	70%	78%
(B_{ICT}, N_{ICT})	79%	74%

verkeer 70% → 78% < BUS, FIETS, TREIN, VRACHTWAGEN

ICT 79% → 74% < MOBIELE TELEFOON, E(-)MAIL



4. Resultaten

Hypothese 4: Rol van de conceptfrequentie

- Kwalitatief de data kennen

Overzicht

1. Inleiding

2. Het Nederlands

3. Methode

4. Resultaten

5. Conclusie

Referenties



5. Conclusie

Bevestiging

- Diachrone convergentie
- Synchrone stratificatie

Opgepast

- ! Belang semantisch veld
- ! Belang van conceptfrequentie

5. Conclusie

Toekomst

Semasiologisch perspectief

- Correlatie onomasiologisch en semasiologisch perspectief

Automatisering: technieken uit de distributionele semantiek

- Vector Space Models (VSMs)
 - Type-based VSMs
 - Token-based VSMs



5. Conclusie

Toekomst - Semasiologisch perspectief

Onomasiologie: gegeven een concept - hoe wordt dit benoemd?

⇒ Semasiologie: gegeven een woord - wat betekent dit?

Bv. AUTOSNELWEG vs. RING

autosnelweg
|
AUTOSNELWEG

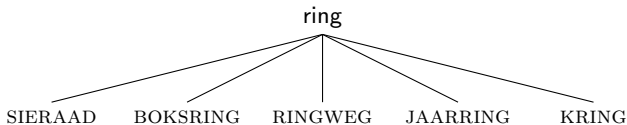
5. Conclusie

Toekomst - Semasiologisch perspectief

Onomasiologie: gegeven een concept - hoe wordt dit benoemd?

⇒ Semasiologie: gegeven een woord - wat betekent dit?

Bv. AUTOSNELWEG vs. RING



5. Conclusie

Toekomst - Semasiologisch perspectief

Case study verkeer:

Onomasiologisch hogere uniformiteit tussen Belgisch Nederlands en Nederlands Nederlands dan semasiologisch, m.a.w. we benoemen concepten eerder hetzelfde dan dat termen dezelfde betekenis hebben.



5. Conclusie

Toekomst - Distributionele semantiek - Vector Space Models

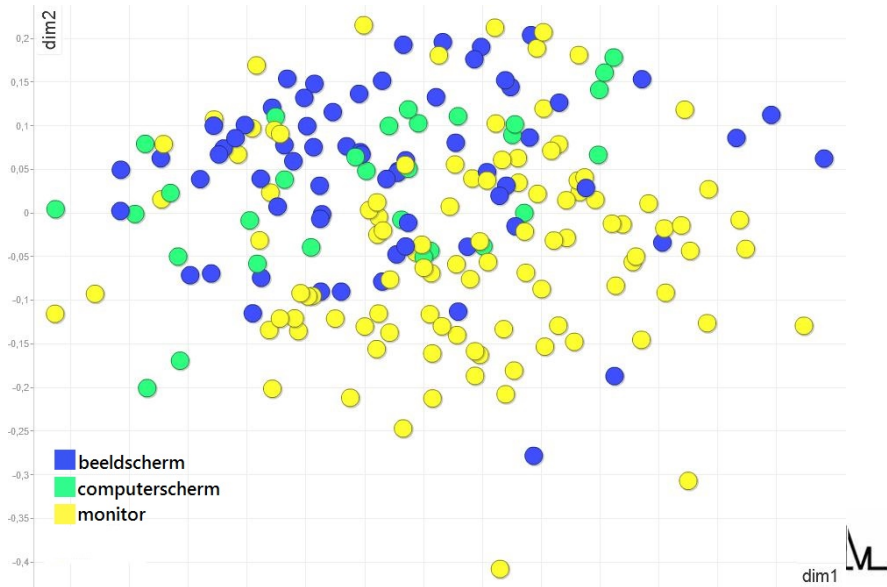
“You shall know a word by the company it keeps” (Firth)

Type-based ~ onomasiologie: synoniemen detecteren

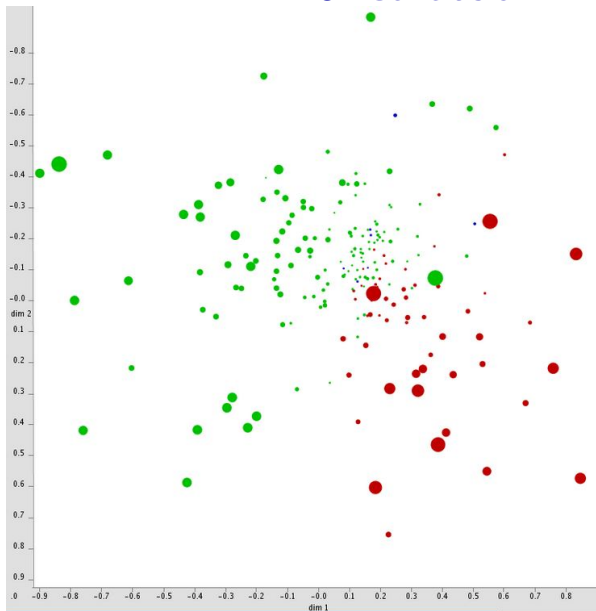
Token-based ~ semasiologie: betekenissen detecteren



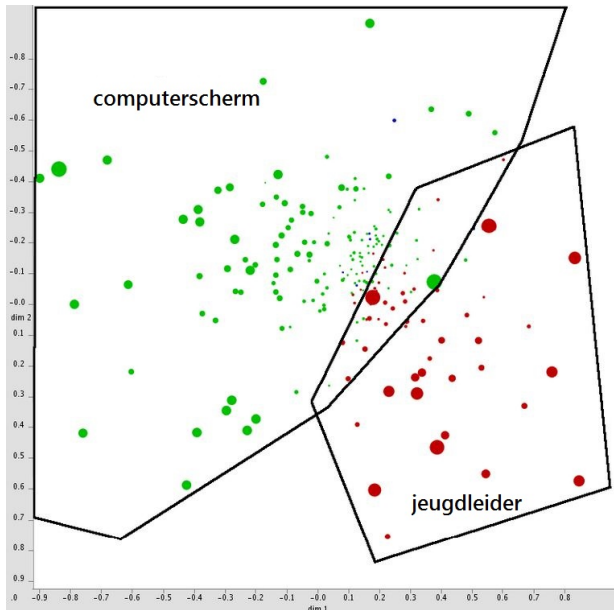
5. Conclusie



5. Conclusie



5. Conclusie



Overzicht

1. Inleiding

2. Het Nederlands

3. Methode

Uniformiteitsmaat

Data

4. Resultaten

5. Conclusie

Referenties



Bedankt! Vragen?



Verdere informatie?

<http://wwling.arts.kuleuven.be/qlvl>
jocelyne.daems@arts.kuleuven.be



Overzicht

1. Inleiding
2. Het Nederlands
3. Methode
4. Resultaten
5. Conclusie

Referenties



Referenties 1/2

- Bouma, Gosse, Gertjan van Noord & Rob Malouf. 2001. "Alpino: wide-coverage computational analysis of Dutch." In: Walter Daelemans, et al. (eds). *Computational Linguistics in the Netherlands 2000*. Amsterdam: Rodopi, 45-59.
- Clyne, Michael. 1992. *Pluricentric languages: differing norms in different nations*. Berlin/New York: Mouton de Gruyter.
- De Cnodder, Tine. 2013. *Convergentie en divergentie in de Nederlandse woordenschat anno 2012*. Unpublished MA thesis. KU Leuven.
- Geeraerts, Dirk, Stefan Grondelaers & Dirk Speelman. 1999. *Convergentie en divergentie in de Nederlandse woordenschat: een onderzoek naar kleding- en voetbaltermen*. Amsterdam: P.J. Meertens-Instituut.
- Grondelaers, Stefan, Dirk Geeraerts, Dirk Speelman and José Tummers. 2001. Lexical standardisation in internet conversations. Comparing Belgium and The Netherlands. In: Fontana, McNally, Turell and Enric Vallduví (eds.), *Proceedings of the First International Conference on Language Variation in Europe*, 90–100. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada, Unitat de Investigació de Variació Lingüística.



Referenties 2/2

- Grondelaers, Stefan, Hilde Van Aken, Dirk Speelman & Dirk Geeraerts. 2001. Inhoudswoorden en preposities als standaardiseringsindicatoren. De diachrone en synchrone status van het Belgische Nederlands. *Nederlandse Taalkunde* 6: 179–202.
- Heylen, Kris, Dirk Speelman & Dirk Geeraerts. 2012. “Looking at word meaning. An interactive visualization of Semantic Vector Spaces for Dutch synsets.” In Miriam Butt, Sheelagh Carpendale, Gerald Penn, Jelena Prokic & Michael Cysouw (eds.), *Proceedings of the EACL-2012 joint workshop of LINGVIS & UNCLH: Visualization of Language Patterns and Uncovering Language History from Multilingual Resources*, 1624. Stroudsburg: Association for Computational Linguistics.
- Speelman, Dirk, Stefan Grondelaers & Dirk Geeraerts. 2003. “Profile-based linguistic uniformity as a generic method for comparing language varieties.” *Computers and the Humanities* 37 (3): 317-337.
- Wielfaert, Thomas, Kris Heylen & Dirk Speelman. 2013. “Visualisations interactives des espaces vectoriels sémantiques pour l'analyse lexicologique.” *Actes de SemDis 2013 : Enjeux actuels de la sémantique distributionnelle*:154-166.

